

**GRAPHICAL WEB BASED TOOL FOR GENERATING QUERY FROM  
STAR SCHEMA**

**MOHAMMED F R ANBAR**

**UNIVERSITI UTARA MALAYSIA  
2009**

# **GRAPHICAL WEB BASED TOOL FOR GENERATING QUERY FROM STAR SCHEMA**

A thesis submitted to the Graduate School, College of Arts and Sciences in partial  
fulfilment of the requirements for the degree Master of Science (IT)

Universiti Utara Malaysia

By

MOHAMMED F.R. ANBAR

(Matric No: 800310)

© MOHAMMED F.R. ANBAR, 2009

All rights reserved

## **PERMISSION TO USE**

In presenting this thesis in partial fulfillment of the requirements for a Master of Science in IT degree from University Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor or, in their absence by the Academic Dean College of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to University Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to

**Dean (Academic) College of Art and Sciences  
University Utara Malaysia  
06010 UUM Sintok  
Kedah Darul Aman.**

## **ABSTRACT**

Novice users have difficulty to generate structured query language from the star schemas because they are not familiar with formulating SQL queries and SQL syntax. This study proposed graphical web based tool to generate queries from star schema and represent the data in tabular or graphical forms which help novice user to formulate SQL query. A prototype for a web based tool to generate the query has been developed using Java Server Pages programming language. The developed tool can facilitate complex query construction which is faced by non-technical and/or novice users. The output of SQL query is presented in tabular and graphical forms which can help users especially top management in better understanding and interpreting query results.

## **ACKNOWLEDGEMENTS**

First, I would like to express my appreciation to Allah, the most merciful and, the most compassionate, who has granted me the ability and willing to start and complete this study. I do pray to His Greatness to inspire and enable me to continue the work for the benefits of humanity.

After that, my most profound thankfulness goes to my supervisor Prof Ku Ruhana Ku-Mahamud for her scientifically proven and creativity encouraging guidance and great support in this study.

Last, I wish to thank my Father, Mother who were always there for me by giving everything they have, my brothers and sisters for their love and support.

Thank you UUM.

Mohammed F.R Anbar

April 12, 2009

# TABLE OF CONTENTS

<b>PERMISSION TO USE.....</b>	<b>I</b>
<b>ABSTRACT .....</b>	<b>II</b>
<b>ACKNOWLEDGEMENT .....</b>	<b>III</b>
<b>TABLE OF CONTENT .....</b>	<b>IV</b>
<b>LIST OF TABLES.....</b>	<b>VI</b>
<b>LIST OF FIGURES.....</b>	<b>VII</b>
<b>LIST OF ABBREVIATIONS.....</b>	<b>VIII</b>
<b>CHAPTER ONE: INTRODUCTION.....</b>	<b>1</b>
1.1 PROBLEM STATEMENT.....	4
1.2 RESEARCH OBJECTIVE.....	5
1.3 SIGNIFICANCE OF THE STUDY.....	5
1.4 SCOPE OF STUDY.....	6
1.5 ORGANIZATION OF THE REPORT.....	6
<b>CHAPTER TWO: LITERATURE REVIEW.....</b>	<b>7</b>
2.1 GRAPHICAL QUERY GENERATION.....	7
2.2 QUERYING IN DATABASE .....	11
2.3 STAR SCHEMA STRUCTURE IN DATA WAREHOUSE.....	15
2.4 STRUCTURED QUERY LANGUAGE FOR FORMULATING QUERY FROM STAR SCHEMA...	17
2.5 WEB APPLICATIONS FOR ENHANCEMENT DATABASE APPLICATION.....	19
2.6 SUMMARY .....	21
<b>CHAPTER THREE : RESEARCH METHODOLOGY.....</b>	<b>22</b>
3.1 PROBLEM ANALYSIS .....	23
3.2 FORMULATION ALGORITHM .....	23
3.3 GRAPHICAL QUERY FORMULATION.....	23
3.4 DOCUMENTATION.....	24
3.5 SUMMARY.....	24
<b>CHAPTER FOUR: ALGORITHM FOR QUERY CONSTRUCTION.....</b>	<b>25</b>
4.1 CONNECTIONAL FRAMEWORK.....	25
4.2 ALGORITHM FOR QUERY CONSTRUCTION.....	27
4.3 THE STRATEGY FOR QUERY RESULT PRESENTATION .....	28
4.3 SUMMARY .....	28

<b>CHAPTER FIVE: SYSTEM ANALYSIS AND DEVELOPMENT .....</b>	<b>29</b>
5.1THE FUNCTIONAL AND NON-FUNCTIONAL REQUIREMENT.....	29
5.2USE CASE DIAGRAM AND SPCIFICATION.....	31
5.3CLASS DIAGRAMS.....	38
5.4SEQUENCE DIAGRAMS.....	40
5.5SYSTEM IMPLEMENTATION.....	44
5.5.1INITIAL INVESTIGATION :.....	44
5.5.2 SYSTEM DESIGN :.....	44
5.5.3CODING AND IMPLEMENTATION:.....	45
5.6 FUNCTIONALITY.....	46
5.7 SUMMARY.....	49
 <b>CHAPTER SIX: PROTOTYPE TESTING.....</b>	<b>50</b>
6.1 SOFTWARE TESTING.....	50
6.2DESIGN OF THE TEST.....	51
6.3 RESULTS.....	51
6.4SUMMARY.....	62
 <b>CHAPTER SEVEN: CONCLUSION.....</b>	<b>63</b>
7.1 CONTRIBUTION OF THE STUDY .....	63
7.2 FUTURE WORK.....	64
REFERENCES.....	65
APPEPDIX A.....	71

# LIST OF TABLES

TABLE 1.1 : REPORTED PREFERENCES FOR GRAPHS OR TEXT .....	2
TABLE 2.1: GENERATING SQL QUERY USING GRAPHICAL TOOL VERSUS GENERATION SQL QUERY USING TEXTUAL TOOL. ....	8
TABLE 2.2: OLTP VERSES OLAP .....	13
TABLE 2.3: DATA WAREHOUSE VERSES OPERATIONAL DATABASE (CHAUDHURI & DAYAL, 1997). ....	14
TABLE 2.4: SIMPLE WEB BASED APPLICATION VERSUS ADVANCE WEB BASED APPLICATION.....	20
TABLE 5.1: SYSTEM FUNCTIONAL REQUIREMENT.....	29
TABLE 5.2:NON- FUNCTIONAL REQUIREMENT FOR SYSTEM.....	31
TABLE 5.3: GENERATE CROSS TAB REPORT USE CASE.....	33
TABLE 5.4: VIEW CROSS TAB REPORT USE CASE.....	34
TABLE 5.5: DELETE CROSS TAB REPORT USE CASE.....	35
TABLE 5.6: DETERMINING CROSS TAB STYLE USE CASE.....	36
TABLE 5.7: UPDATE CROSS TAB STYLE USE CASE.....	37
TABLE 5.8: DELETE CROSS TAB STYLE USE CASE.....	38
TABLE 6.1: TEST CASE ADMINISTRATOR LOGIN FUNCTIONALITY.....	51
TABLE 6.2: TEST CASE GENERATES CROSS TAB REPORT FUNCTIONALITY.....	52
TABLE 6.3: TEST CASE VIEW CROSS TAB REPORT FUNCTIONALITY .....	54
TABLE 6.4: CASE DELETE CROSS TAB REPORT FUNCTIONALITY .....	54
TABLE 6.5: TEST CASE VIEW CROSS TAB REPORT FUNCTIONALITY .....	55
TABLE 6.6: TEST CASE DELETE CROSS TAB REPORT FUNCTIONALITY .....	55
TABLE 6.7: TEST CASE DELETE CROSS TAB STYLE FUNCTIONALITY .....	56
TABLE 6.8: TEST CASE VIEW CROSS TAB REPORT STYLE FUNCTIONALITY .....	56
TABLE 6.9: TEST CASE GENERATE VERTICAL BAR REPORT FUNCTIONALITY .....	57
TABLE 6.10: TEST CASE VIEW VERTICAL BAR REPORT FUNCTIONALITY .....	58
TABLE 6.11: TEST CASE DELETE VERTICAL BAR REPORT FUNCTIONALITY.....	59
TABLE 6.12: TABLE 6.12: TEST CASE GENERATE PI CHART REPORT FUNCTIONALITY ...	59
TABLE 6.13: TEST CASE VIEW PI CHART REPORT FUNCTIONALITY .....	60
TABLE 6.14: TEST CASE DELETE VERTICAL BAR REPORT FUNCTIONALITY.....	61
TABLE 6.15: TEST CASE USER LOGIN FUNCTIONALITY .....	61



## LIST OF FIGURES

User friendly interface represents visual query that results in viewing operation.	9
The query refinement cycle.....	10
A multidimensional Model.....	14
Star Schema components.....	16
Star Schema representation.....	17
Research methodology stages.....	22
Prototyping System Development Methodology.....	24
The Existing Framework.....	25
Cross tab form of SQL output presentation.....	26
Vertical bar chart of SQL output presentation.....	26
Star schema structure in data ware house.....	27
Use case digram for web based tool.....	32
Class Diagram for the proposed System.....	39
Sequence Diagram for user to generate crosstab report use case.....	40
Sequence Diagram for user to view crosstab report use case.....	41
Sequence Diagram for user to delete crosstab report use case.....	41
Sequence Diagram for administrator to update crosstab report style use case...	42
Sequence Diagram for administrator to determine crosstab report style use case.	43
Sequence Diagram for administrator to delete crosstab report style use case.....	43
Prototyping System Development Methodology.....	44
Main screen for the System .....	46
Main page for administrator.....	47
Main page for determining cross tab report style.....	47
Presenting output of SQL in cross tab report form.....	48
Presenting output of SQL in vertical bar chart form.....	48
Presenting output of SQL in PI chart form.....	49

## LIST OF ABBREVIATIONS

SQL	Structured query language
OLAP	On-line analytical processing
OLTP	Online transaction processing
HTML	HyperText Markup Language
IDE	Integrated Development Environment
JSP	Java Server Pages
DW	Data warehouse
NL	Natural language
SQUARE	Specifying Queries as Relational Expressions
TRC	Tuple relational calculus
DDL	Data definition language
DML	Manipulation language
DBMS	Database management system
RDBMS	Relational Database Management system
RUP	Rational Unified Process
UML	Unified Modeling Language
WWW	WORLD WIDE WEB
HTTP	Hyper Text Transfer Protocol

<b>NO</b>	<b>TITLE OF THE FIGURE</b>	<b>PAGE</b>
2.1	User friendly interface represents visual query that results in viewing operation	9
2.2	The query refinement cycle	10
2.3	A multidimensional Model	14
2.4	Star Schema components	16
2.5	Star Schema representation	17
3.1	Research methodology stages	23
3.2	Prototyping System Development Methodology	25
4.1	The Existing Framework	27
4.2	cross tab form of SQL output presentation	28
4.3	vertical bar chart of SQL output presentation	28
4.4	star schema structure in data ware house	29
5.1	Use case digram for web based tool	34
5.2	Class Diagram for the proposed System	40
5.3	Sequence Diagram for user to generate crosstab report use case	41
5.4	Sequence Diagram for user to view crosstab report use case	41
5.5	Sequence Diagram for user to delete crosstab report use case	42
5.6	Sequence Diagram for administrator to update crosstab report style use case	43
5.7	Sequence Diagram for administrator to determine crosstab report style use case	44
5.8	Sequence Diagram for administrator to delete crosstab report style use case	44
5.9	Prototyping System Development Methodology	45
5.10	Main screen for system	47
5.11	Main page for administrator	48
5.12	Main page for determining cross tab report style	48
5.13	presenting output of SQL in cross tab report form	49
5.14	presenting output of SQL in vertical bar chart form	50
5.15	presenting output of SQL in PI chart form	51

## **CHAPTER ONE**

### **INTRODUCTION**

Database applications which are deployed in many corporations for the purpose of storing and retrieving data in a structured way has become ubiquitous and indispensable. Many transactions are being performed by users daily. These transactions represent the questions asked by users in the natural language which are formulated as structured query language (SQL) to deal with database. The ways of formulating queries depends on user knowledge and experience. Ultimately, the output of queries represents the answers to user questions.

A query language is a specialized language in which a user requests information from a database. These are typically of a higher-level than programming languages. They are classified in to two types, procedural and non procedural. Procedural is where the user instructs the system to perform a sequence of operations on the database. That will compute the desired information. Nonprocedural is where the user specifies the information desired without giving a procedure in obtaining the information. A complete query language also contains facilities to insert and delete topples as well as to modify parts of the existing topples.

Most of data set which represents the outputs of SQL query are presented in the form that users take a lot of time in browsing the data. This makes the users uneasy

The contents of  
the thesis is for  
internal user  
only

## References:

- Agarwal, D., Barman, D., Gunopulos, D., Young, N. E., Korn, F., & Srivastava, D. (2007). Efficient and effective explanation of change in hierarchical summaries. *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*.
- Baresi, L., & Morasca, S. (2007). Three empirical studies on estimating the design effort of Web applications. *ACM Transactions on Software Engineering and Methodology*, 16 (4), 40-45
- Becker, K., Cardoso, M. O., Nichele, C. M., & Frighetto, M. (2000). Mail-by-example: a visual query interface for email management. *Proceedings of the working conference on Advanced visual interfaces*. 280-281.
- Bonifati, A., Cattaneo, F., Ceri, S., Fuggetta, A., & Paraboschi, S. (2001). Designing data marts for data warehouses. *ACM Transactions on Software Engineering and Methodology*, 10(4), 452-483.
- Borgman, C. L. (1996). Why are online catalogs still hard to use? *Journal of the American Society for Information Science*, 47(7), 493-503.
- Brambilla, M., Ceri, S., Fraternali, P., & Manolescu, I. (2006). Process modeling in Web applications. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 15(4), 360-409.
- Cao, B., & Badia, A. (2007). SQL query optimization through nested relational algebra. *ACM Transactions on Database Systems (TODS)*, 32(3)

- Chapple, M. (2000). Data Definition Language. Retrieved February 11<sup>th</sup>, 2009 From [http://databases.about.com/od/sql/a/sqlfundamentals\\_2.htm](http://databases.about.com/od/sql/a/sqlfundamentals_2.htm)
- Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *ACM SIGMOD Record*, 26(1), 65-74.
- Chen, Z., & Li, T. (2007). Addressing diverse user preferences in SQL-query-result navigation. *Proceedings of the 2007 ACM SIGMOD international conference on management of data*. 641-652.
- Cl, B. (1999). The users mental model of an information retrieval system: an experiment on a prototype online catalog. *International Journal of Human-Computer Studies*, 51(2), 435-452.
- Conallen, J. (2000). Modeling Web Applications with UML. *White paper, Conallen Inc.*, <http://www.conallen.com/whitepapers/webapps/ModelingWebApplications.htm>, March 1999.
- Conallen, J. (2002). *Modeling Web Applications with UML* (2<sup>nd</sup> ed.). USA: Addison-Wesley.
- Connolly, T. M., & Begg, C. E. (2005). *Database Systems: A Practical Approach to Design, Implementation, and Management*: USA: Addison Wesley Publishing Company.
- Deshmukh, H., Malavia, J., & Carter, w. J. (2003). *SCWCD exam study kit : Java web component developer certification*. Greenwich: Manning Publications.
- France, R., & Kobryn, C. (2001). UML for software engineers. *Proceedings of the 23rd International Conference on Software Engineering*. 705-706

- Ginige, A., & Murugesan, S. (2001). Web engineering: an introduction. *IEEE Multimedia*, 8(1), 14-18.
- Hao, J., & Mendes, E. (2006). Usage-based statistical testing of web applications. *Proceedings of the 6th international conference on Web engineering*. 17 - 24
- Huo, J. (2008). *KMVQL: a visual query interface based on Karnaugh map*. Proceedings of the working conference on Advanced visual interfaces.
- Inmon, B., & Kelly, C. (1994). The Twelve Rule of Data warehouse for a Client/Server World. *Data Managment Review* 4(5), 6-16.
- Jarke, M. (2003). *Fundamentals of Data Warehouses*. New York: Springer-Verlag.
- Jones, S. (1998). Graphical query specification and dynamic result previews for a digital library. *Proceedings of the 11th annual ACM symposium on User interface software and technology*.
- Jung, H., & Lee, G. G. (2002). Multilingual question answering with high portability on relational databases. *COLING-02: proceedings of the 2002 conference on multilingual summarization and question answering - Volume 19*
- Kaner, C. (2003). *What is a Good Test Case?*
- Kappel, G., Pröll, B., Reich, S., & Retschitzegger, W. (2006). *Web Engineering-The Discipline of Systematic Development of Web Applications*. USA: John Wiley & Sons, Inc.
- Kern, J., & Garrett, C. (2003). Effective Sequence Diagram Generation. *Borland* ([www.borland.com/together/white\\_papers](http://www.borland.com/together/white_papers)).
- Kim, W. (1982). On optimizing an SQL-like nested query. *ACM Transactions on Database Systems (TODS)*, 7(3), 443-469.



- Krippendorff, M., & Song, I. Y. (1997). The translation of star schema into entity-relationship diagrams. *Proceedings of the 8th International Workshop on Database and Expert Systems Applications*.
- Law, A. S., Freer, Y., Hunter, J., Logie, R. H., McIntosh, N., & Quinn, J. (2005). A Comparison of Graphical and Textual Presentations of Time Series Data to Support Medical Decision Making in the Neonatal Intensive Care Unit. *Journal of Clinical Monitoring and Computing*, 19(3), 183-194.
- Malinowski, E., & Zimányi, E. (2006). Hierarchies in a multidimensional model: From conceptual modeling to logical representation. *Data & Knowledge Engineering*, 59(2), 348-377.
- Martin, R. C. (1997). UML Tutorial: Part 1--Class Diagrams. *Engineering Notebook Column, C++ Report*.
- Nourie, D. (2006). Java Technologies for Web Applications. Retrieved January 13<sup>th</sup>, 2009 From [http://java.sun.com/developer/technicalArticles/tools/webapps\\_1/](http://java.sun.com/developer/technicalArticles/tools/webapps_1/).
- Pokorny, J. (1998). Data Warehouses: a Modelling Perspective. In: Evolution and Challenges in System Development. *Proceeding of 7th International Conference on Information Systems*.
- Ramakrishnan, R., & Gehrke, J. (2003). *Database Management Systems* (3 ed.). New York: McGraw-Hill Professional.
- Rob, P., & Coronel, C. (2008). *Database Systems: Design, Implementation, and Management*. USA: Thomson Course Technology.

- Rowen, W., Song, I. Y., Medsker, C., Arynth, I., & Ewen, E. (2001). An Analysis of Many-to-Many Relationships Between Fact and Dimension Tables in Dimensional Modeling. *Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW'2001) Interlaken, Switzerland, June 4, 2001.*
- Sibley, E. H., & Taylor, R. W. (1973). A data definition and mapping language. *Communications of the ACM*, 16(12), 750-759.
- Stoffel, K., Davis, J. D., Rottman, G., Saltz, J., Dick, J., Merz, W., et al. (1998). A graphical tool for ad hoc query generation. *Proceedings of the AMIA(American Medical Informatics Association) Symposium*. 503-7.
- Nishith. (2005). Database vs. Data Warehouse. Retrieved February 2<sup>nd</sup>, 2009, From <http://opensourceanalytics.com/2005/11/02/database-vs-data-warehouse/>
- Weininger, A. (2002). Efficient execution of joins in a star schema. *Proceedings of the ACM SIGMOD international conference on Management of data.*